

DRAFT not for Dissemination



ESSI Expert Help Task “A Comprehensive School/District Data Model”

***Phase One:
A Guiding and Stakeholder Convening Draft***

October 7th, 2005



Comprehensive pK-12 Education Data Model

Table of Contents

I.	Introduction	3
II.	pK12 Data Models	4
	a. What is a Data Model?	
	b. Data Models for Various Audiences	
	c. Past attempts at Data Models	
	d. Sample Data Models	
	e. Examples of National Efforts	
	f. Examples of State Efforts	
	g. Examples of Local Efforts	
III.	Project Scope and Parameters	14
	a. Identifying the Target Audience for a Comprehensive Education Data Model	
	b. Identifying the Targeted Levels of the Model	
	c. Identifying Levels of Detail in the Model	
	d. Factoring the Categories and Sections	
	e. Developing a Common Vocabulary	
	f. Identifying Essential Data Model Components	
IV.	Project Findings	17
	a. Summary of Data Model Comparisons	
	1. Preliminary Data Model Sections	
	2. Organization Domains	
	3. Person Domains	
	b. The Developed Domain Framework	
V.	Suggested Next Steps	23
VI.	Conclusions	24
	Appendices	25
	a. Data Model Comparison Crosswalk	
	b. Data Handbook	
	c. Other Defined Handbook Entity Terms	



Comprehensive pK-12 Education Data Model

I. Introduction

The passage of the *No Child Left Behind Act* (NCLB) increased the demand for more immediate access to high-quality education data at all levels of pK-12 education enterprise. The bulk of information needed to respond to this and other federal reporting requirements is initially captured in local schools and district offices to support local operations and decision-making. A subset of this data is then subsequently moved through state information systems to the United State Department of Education (USED). The dependence of federal, state and local decision makers on the quality and availability of education data has never been higher.

The data points necessary to make these determinations come from sources across the pK-12 enterprise including student enrollment by socio-economic status (usually student eligibility for free or reduced price lunch), student performance on standardized tests disaggregated by race and ethnicity class schedules, staff qualifications (teaching credentials), and teaching assignments.

This increase in the use of accountability models and diagnostic analyses to improve student achievement to fulfill extensive reporting requirements is leading districts and states to design, develop, implement, and apply sophisticated data systems. The architecture and content of these data systems define the breadth and depth of research and diagnostic analyses each system is able to support. The content of state data systems is dependent on the data collected by district data systems which, thereby affects the state's ability to aggregate and use data for analysis and planning. The content and architecture of state data systems, in turn, limit the systems' analytical capabilities and, therefore, their ability to support instructional improvement and help close student achievement gaps. The content and structure of data systems also affect the states' ability to share data on students who move across state boundaries.

The only way to meet these demands is through the development of a comprehensive and dynamic common pK-12 data model. To date, the closest access to data models that schools and states have is in proprietary models developed by vendors and implemented in their software applications. With the majority of districts (more than 10,000) in the US having fewer than 2,500 students, many districts cannot afford these data solutions or they must enter into consortium purchases which do not allow them to tailor their data utilization to the needs of their educational stakeholders. Most states and school districts cannot make the necessary financial investment or do not have the technical expertise to develop such comprehensive data models. The closest available public examples available have been developed to ensure conformity to mandated reporting information such as USED's Performance Based Data Management Initiative (PBDMI) and state reporting requirements including "silo" requests from the various offices within state agencies to meet some specific reporting need. None of these public examples extend to the classroom level to describe that vital component in the teaching and learning process (student progress data) and help in the efficient allocation of limited resources at the local level.

The development of such a model must include a thorough evaluation of data needs at the local, regional, state, and federal levels. Many districts and states want to design data



Comprehensive pK-12 Education Data Model

warehousing and data collection sites with the purpose of greater tracking of longitudinal data. They are looking for guidance on what data should be included in their systems. Without a comprehensive data model, they will continue with existing models being developed that may not meet the needs for comprehensive collections, and will not impact the educational learning environment to the greatest extent possible.

A comprehensive data model must be completed as soon as possible to provide maximum support for current state and local data system developments. This work cannot be done without first examining existing data models at the national, state, and local levels and reconciling their common components into a high-level framework. That examination and reconciliation is the work of this task order and this document is the first step in that strategy.

This document assembles past and current attempts to produce a data model, including a catalog and examples (where available) of existing data model work.

II. pK12 Data Models

What is a Data Model?

A data model is a conceptual representation of the data structures (tables) that are used in expressing and communicating business requirements. Done correctly, models visually represent the nature of data, the business rules governing the data, and how the data will be organized in a database. The goal of a data model is to make sure that all provided data objects are completely and accurately represented. A data model can also be used to communicate within the business and across business domains - as seen in education.

A data model also may provide a method of visualizing the informational needs of a system. It typically takes the form of an ERD (Entity Relationship Diagram) and almost always must include a data dictionary of some kind to be a complete. Three basic types of data models are used:

Conceptual Data Model (CDM) - A CDM represents the overall logical structure of the sets of data, independent of any software or data storage structure. A conceptual model usually contains data objects not yet implemented in the physical databases. It gives a formal representation and factoring of the data sets needed to run an enterprise or a business activity.

Logical Data Model (LDM) - A LDM fills in the gap between a conceptual and physical data model (PDM). CDMs are completely devoid of database-level information while PDMs are specific to a certain database management system (DBMS). The LDM contains those characteristics of the data structure that are generic in nature such as indices and foreign keys. A LDM stores those characteristics without adding anything specific to a single DBMS.

Physical Data Model (PDM) - The PDM specifies the physical implementation of the database. With the PDM, one considers the details of actual physical implementation and the specific data elements correlated to fields, dimensions or facts in the actual data base. It takes into account both software or data storage structures. People often modify the PDM to suit their performance or physical constraints. As a result, it can be difficult to backward engineer



Comprehensive pK-12 Education Data Model

from a PDM; the PDM almost always is no longer organized in logical business-process-recognizable data groupings.

Data Models for Various Audiences

Data models take various forms depending on the needs of various audiences for contextualized information. Conceptual models tend to lend themselves to a higher level stakeholder audience such as administrators. This type of model provides a “high level” overview of the flow of data without getting into the granular details. This model also paints the “big picture” of the data needs, in this case across the pK-12 realm.

Logical data models are more granular in detail and are useful to audiences who want to understand the relationships between the data elements and the desired flow of data. While this type of model still captures the conceptual outline of the data, the logical or expected flow of data is exposed in greater details. This type of model is typically used by data administrators and managers.

The physical data model is one that is the hardest to display and maintain. This type of model is often used by the software developer and technology audience. It is ever-changing and evolving as systems are added and removed from the model. This type of model is said to “never be finished.”

Past Attempts at Data Modeling

A complete picture of the pK-12 environment would not only facilitate the visualization of data transfer between applications within schools, but also conceptualize how to streamline reporting from the school, to district, regional, state, and ultimately to the federal level. In addition, this data model would facilitate the development of applications designed to provide administrators and educators with the tools they need to both collect and analyze the achievement and related data necessary to improve student performance.

While there have been efforts to describe these relationships, these models tended to be narrowly focused at the local (report generation), state (reporting for accountability/funding) and national (PBDMI/EDEN) levels. Each effort has focused on a particular set of data identified to fulfill some mandated/operational task. The majority of work done around the development of comprehensive pK-12 data models can be seen within the vendor community in the development of data warehouses and/or student information systems to be sold at the local or state level. These product lines have been comprehensively developed, although most do not have a clear understanding of educational business cases and relationships. They also receive ongoing maintenance to promote company proprietary business advantages.

This project has been designed to gather examples of past efforts, develop a high level representation of the extent of work, and provide a “gaps and overlap” analysis of data currently utilized in the pK-12 space. This representation will include various national and state “data models” which takes into account representations of the data needed for education planning, operations, and reporting.



Comprehensive pK-12 Education Data Model

One of the conclusions of Phase I of this project is also one of its assumed starting points: most districts do not have the money, technical resources, or time to build a fully-developed data model. Instead they build their data structures piecemeal out of what they get from each vendor or as each project is completed. Thus there is little or no uniformity or standardization, upgrades become difficult to complete and solutions must include multiple vendors with different requirements. While that situation may not change, the presence of a definitive comprehensive example of a data model, can serve as a starting point to significantly improve the design, implementation, and utility of education data systems.

There is currently no comprehensive data model at the national level. There are numerous data dictionaries for various audiences: the National Center for Education Statistics (NCES) data handbooks, the USED PBDM/EDEN data structures, and the Schools Interoperability Framework Association (SIFA) data model. They, however, provide little guidance on the relationships and interconnectedness between those structures and data sets. At the state and district level almost all models are physical models and representations of their database structure.

Let us speak briefly about what this project WILL NOT do and what it WILL do:

- The first phase WILL NOT create a full national data model – it WILL create the framework upon which that data model can be created in a possible Phase II.
- It WILL NOT be the “right” way to design any organization’s data structure. Rather it WILL provide guidance and templates for building the relationships and data hierarchies that represent the most common shared data requirements of pK-12 organizations in the US.
- It WILL NOT be a template that can just be plugged in and played as a whole. However, as an organization starts to design its data model, it will provide the foundation upon which that organization can make decisions and allow it to do so without re-inventing the wheel. The organization can then focus on making the model match its particular business and educational needs - or it may only use those modules of the model that the organization finds useful.

Reasons supporting the development of a true pK-12 data model would be that:

- A full model WILL provide the core data structures to allow SIF to be made operational.
- A full model WILL provide the core data structures needed for the shared federal data reporting burden to be addressed.
- A full model WILL provide a common vocabulary and framework for entities to have shared conversations about curriculum, data exchange and standards.
- A full model WILL provide guidance on how data on classroom activities and programs relate to other components of the data system. This is an area of critical need in promoting effective decision-making regarding student achievement.



Comprehensive pK-12 Education Data Model

Sample Data Models

While various “data models” exist in different forms and fashions, the key is that they contain the relevant data and relationships needed for the task outlined. However, as mentioned above, these models do not provide all of the components needed to be comprehensive. Presented here are some examples of “data models” that were used in determining the need for a comprehensive data model. These models were chosen for a variety of reasons.

- The NCES, USED and SIFA data models were selected because they come the closest representing a national pk-12 data model.
- The states providing models were self-selected. There are few states that had true models, and these states are the ones with the most developed data models OR they seemed to be examples of the current status of state modeling.
- Most districts do not have a data model in a form that is easily shared with others. Many do not have a true understanding of what is included in a data model. Those districts whose models we utilized in this analysis were those that provided data models that could be compared and contrasted to the other models.

Examples of National Efforts

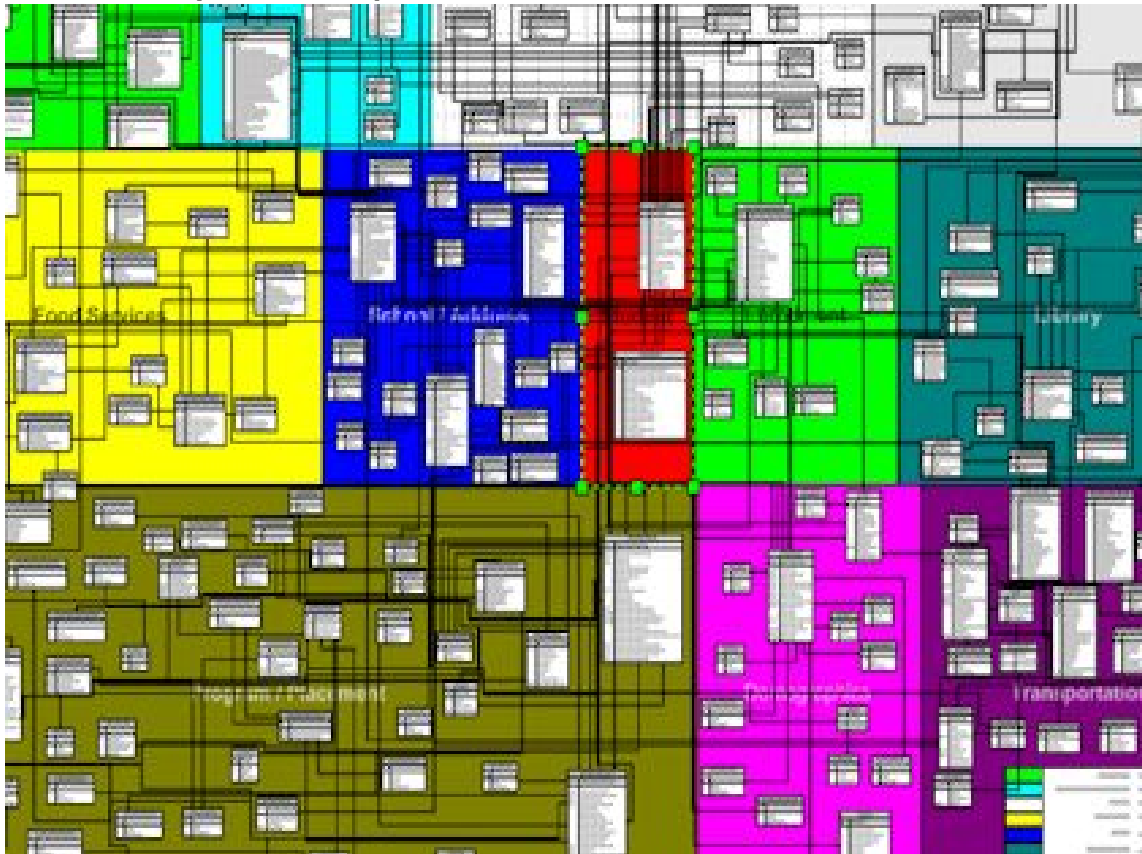
1. Schools Interoperability Framework Association <http://www.sifinfo.org/>

The Schools Interoperability Framework Association (SIFA) is a non-profit membership organization made up of over 300 schools, districts, states, the U.S. Department of Education, software vendors and consultants. These entities work together to collectively define the rules and regulations for educational software data interoperability. SIFA enables diverse applications to interact and share data efficiently, reliably and securely regardless of the platform hosting the applications. It has united these education technology end users and providers in an unprecedented effort to give teachers more time to do what they do best: teach.

This graphical view of the SIF Association data model below illustrates the weakness of using graphical notation in some situations. While space here does not allow for its full view and utilization, each color here represents different software applications used at the local level (i.e. food service, transportation, data warehouse, etc) and each box is a data object used in that application and shared with others. For this model to be useful it would need to be generalized so that the content field is less cluttered. However developers will often have a wall-size version of this kind of data model hanging in their office so they can “walk through” data flows.



Comprehensive pK-12 Education Data Model



2. PBDMI/EDEN Project (USED)

<https://eden.ed.gov/EDENPortal/>

The Performance-Based Data Management Initiative (PBDMI) represents a significant effort by the U.S. Department of Education (USED) to consolidate data collection about pk-12 education and reduce the reporting burden on state education agencies (SEAs). The result is the Education Data Exchange Network (EDEN), which collects data throughout the school year as data become available in SEAs. The data that are collected cover program areas such as No Child Left Behind, Title I, Special Education, Vocational Education, Safe and Drug Free Schools, Migrant Education and education for Limited English Proficient (LEP) students. With the exception of institutional identification and descriptors/statuses of schools, districts, intermediate education units, and state education agencies, most data are aggregate counts. Efforts are made to standardize data elements across programs so that SEAs are not forced to report conflicting or overlapping data. A crosswalk has been done between SIFA data objects and elements and the EDEN data elements to identify where EDEN data elements are not currently addressed by SIFA (example below).



Comprehensive pK-12 Education Data Model

EDEN ID	EDEN Data Element Name	EDEN File Name	EDEN Disaggregation	EDEN Code Set	EDEN Level	SIF Objects/Elements	SIF Missing Codes	SIF Data Elements Needed
1	LEA Identifier (NCES)	Directory		National Center for Education Statistics (NCES) code set: The LEA identification number is made up of a two-digit state code followed by a five-digit local education agency (LEA) code.	LEA	LEAInfo (NCESId)		
					SCH	SchoolInfo (NCESId)		
4	LEA Identifier (State)	Directory		String	LEA	LEAInfo (StatePrId)		
					SCH	SchoolInfo (StatePrId)		

3. National Center for Educational Statistics (NCES) Handbook Information

<http://nces.ed.gov/programs/handbook/index.asp>

The NCES Online Handbook contains a listing of data elements useful to education entities seeking to design and implement information systems. The handbook contains a hierarchical listing of all the data elements that might be kept in an automated record-keeping system about students, staff, or education entities. Consideration was given to including the data needed for federal and state reporting. The intent of the handbook is not to provide a model for data collection and maintenance; rather it is to provide a list of “best practice” definitions of commonly used data elements and comprehensive code sets where possible.

The student and staff sections were developed first. They are independent, but contain identical data elements where logical. The education organization sections were developed subsequently as a means of describing the service providers in education. While it may appear that the handbook represents a “model” of education data, there has been little effort made to show the relationships among the levels and contents of the data elements, categories, entities, sections, and domains. In the past, the handbooks were static for a number of years. Now, there is a procedure for updating the handbook on an annual basis as new data elements are identified and needed.

This is a mature example of an effective data modeling tool. Although graphical interfaces can often give the “feel” or the high-level “view” of the data, to work with a data model one needs to be able to effectively compare its data elements to data elements and relationships from other sources. As a result many of the most effective data models are primarily housed in a spreadsheet format. Still, this “model” does not indicate relationships among all of the components.



Comprehensive pK-12 Education Data Model

4. NCES Common Core Data (CCD)

<http://nces.ed.gov/ccd/>

The Common Core of Data (CCD) is a collection of basic statistical information on the children receiving free public preschool, elementary and secondary education in the U.S. The CCD contains three basic types of data: identifying and descriptive information about the entities within the education system, basic statistics about students and staff, and fiscal data. The National Center for Education Statistics collects these data annually from state education agencies. The database serves as an official listing of public elementary and secondary schools and education agencies in the US and provides basic descriptions about the schools and agencies. It is used for identifying representative samples for further research. This “model” is very limited, and does not show relationships among any of the components.

Examples of State Efforts

1. California Department of Education

<http://sif.edreform.net/resource/11051>

This is an example of a narrative data model. It is documented in a structured documentation format. In this case it serves a particular purpose for the Department and may or may not be considered into the large logical data model for LEA’s to align.

Data Subject Name	Definition
Degree	Any type of degree or honorary degree or title of any designation, mark, appellation, series of letters or words such as, but not limited to, associate, bachelor, master, doctor, or fellow which signifies, purports, or is generally taken to signify satisfactory completion of the requirements of an academic, educational, technological, or professional program of study beyond the secondary school level or is an honorary title conferred for recognition of some meritorious achievement.

Data Characteristic Name	Definition
Degree Major Indicator	A primary field of study in pursuit of a bachelor or associate degree, implying that the individual has substantial knowledge of the academic discipline or subject area.

Data Characteristic Name	Definition
Degree Minor Indicator	A secondary field of study in pursuit of a bachelor or associate degree, implying that the individual has substantial knowledge of the academic.

Data Characteristic Name	Definition
Degree Conferring Date	The month, day, and year on which an individual received a degree.

2. Virginia Department of Education

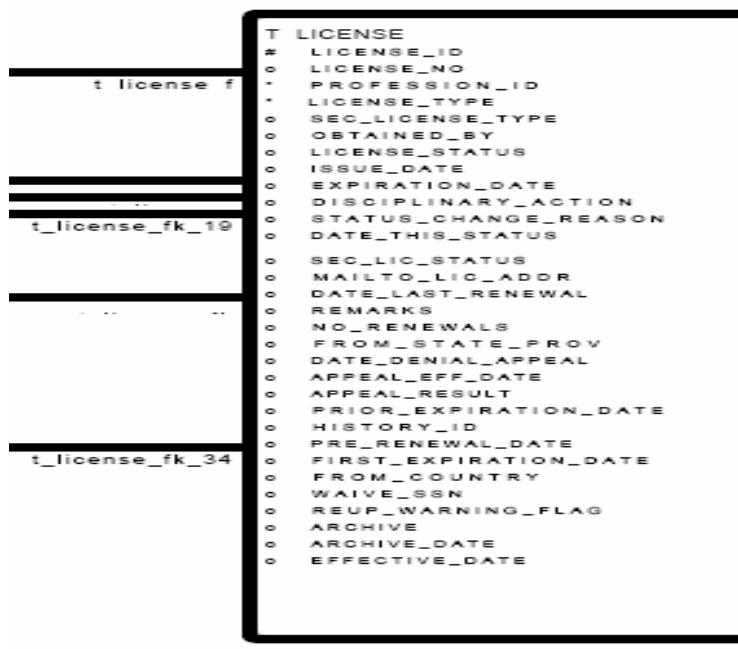
<http://www.pen.k12.va.us/VDOE/Publications/ERD/erd.html>

This format is the more common graphical notation which includes identifying the container object at the top and then listing the elements underneath it. Lines go from the object to similar objects connoting the relationship of this object to the others. Again while it may be a



Comprehensive pK-12 Education Data Model

component within the larger data needs at the state level, this model outlines requirements for the state's data management - and not the local level.



3. Maine Department of Education

<http://www.maine.gov/education/datalist.htm>

This is another example of a data model in the form of structured narrative documentation. Again while serving a function for definition requirements, it does not provide guidance for local data utilization.

Function ID	1000	
Function	Title	Description
Instruction	Regular Instruction	Instruction includes the activities dealing directly with the interaction between teachers and students. Teaching may be provided for students in a school classroom, in another location such as a home or hospital, and in other learning situations. Instruction may also be provided through some other approved medium such as television, radio, telephone, and correspondence that is delivered inside or outside the classroom or in other teacher-student settings. Included here are the activities of aides or classroom assistants of any type who assist in the instructional process.

Function ID	2000	
Function	Title	Description
Support Services	Support Services	Support services are those services which provide administrative, technical (such as guidance and health), and logistical support to facilitate and enhance instruction. Support services exist as adjuncts for the fulfillment of the objectives of instruction, community services and enterprise programs, rather than as entities within themselves.

Function ID	2100	
Function	Title	Description
Support Services	Support Services - Student	Activities designed to assess and improve the well-being of students and to supplement the teaching process.

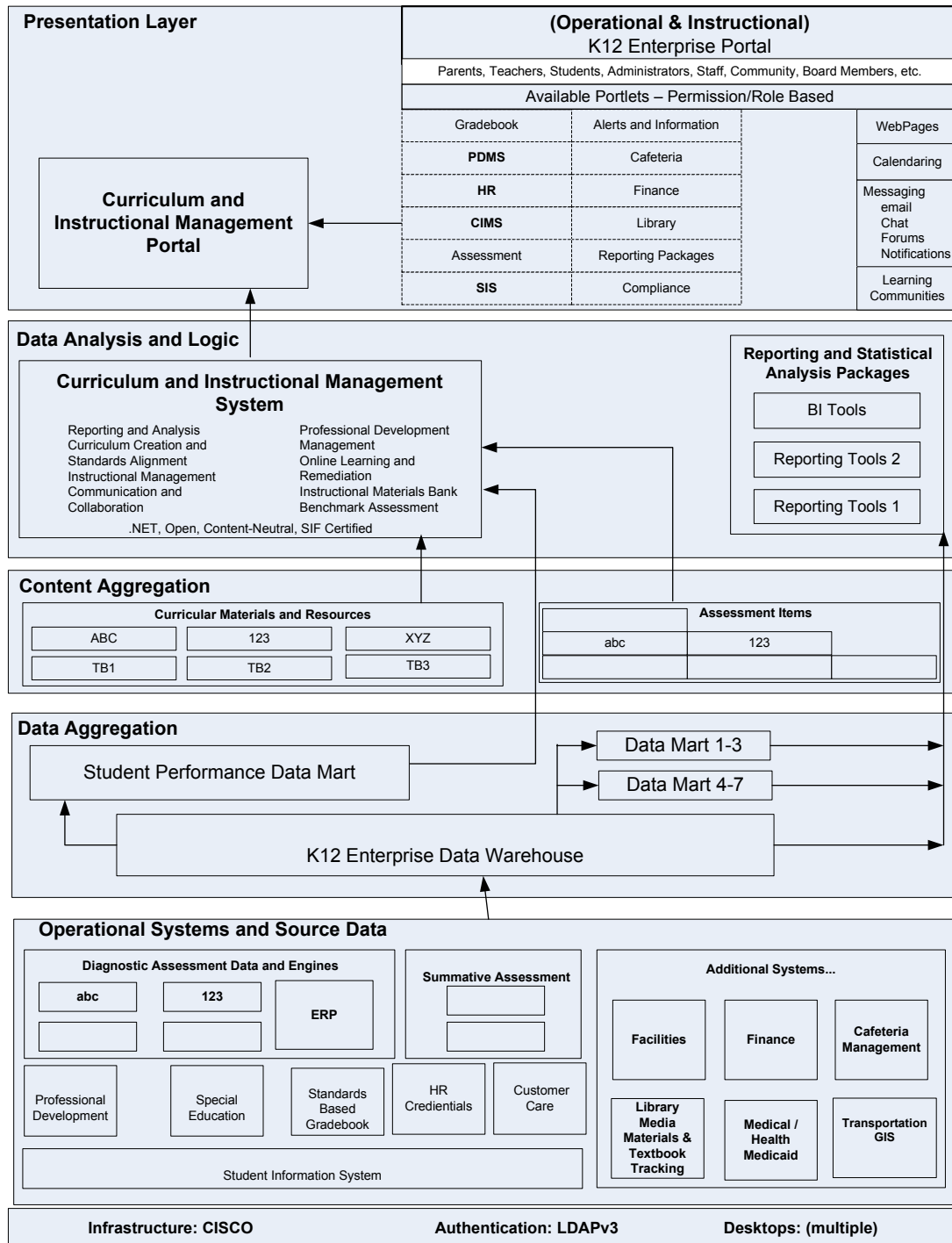


Comprehensive pK-12 Education Data Model

Examples of Local Efforts

1. Washington DC Public Schools

This is an example of a combination logical and physical data model. It represents of how a district uses information in the context of data flowing between applications but it excludes larger data needs and utilization.

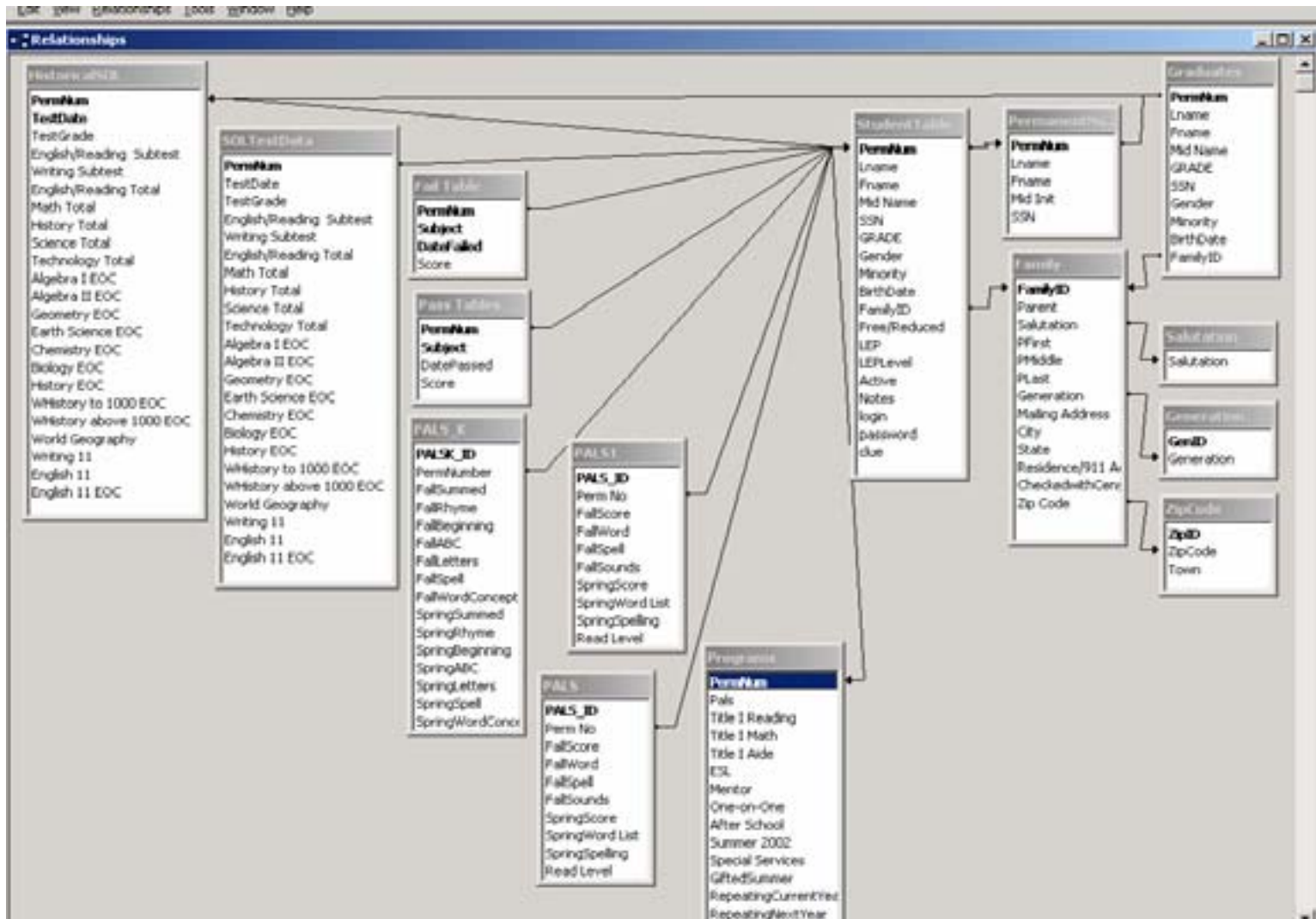




Comprehensive pK-12 Education Data Model

2. Richmond County Schools, Warsaw, Virginia

This is another data model that is representative of what many districts have today - application specific. Basic tables are used to identify the reports and the data elements contained in them.





Comprehensive pK-12 Education Data Model

III. Project Scope and Parameters

This project was conducted to determine how best to provide guidance and best practices for states and districts putting together their enterprise data structures. National Center for Educational Statistics (NCES) has shown leadership and expressed the desire that this project identify the need and scope of work for a possible follow-on project to generate a comprehensive pK-12 education data model. This document discusses the need for a comprehensive data model and provides a first attempt at categorizing how pK-12 data will be sorted and represented in the NCES state and district data models - or a high level conceptualization of the data model.

In this project, a critical component is understanding what makes up a data model and what type of data model is really needed by the education community. To accomplish this, a series of activities were conducted:

- An advisory committee made up of state and local data systems administrators, SIFA data experts, and others knowledgeable about state and local education data met on two occasions. They worked virtually to discuss topics such as target audiences both for this document and for a comprehensive education data model, to provide guidance on model reviews, and to provide input into the recommendations contained in this document.
- Sample data models were reviewed to identify essential components and initial indications of data relationships. This “gap analysis” provided input into what components should be called, helped to determine what the ideal product should look like, and justified the need for a comprehensive education data model.

The information that follows describes the expert panel’s activities, decisions, and recommendations. It also represents the beginning of the process needed to generate a comprehensive education data model, including an understanding of the current landscape of state and district data models and the additional activities that would be needed to accomplish the actual construction of the model in a second project phase.

Identifying the Target Audience for a Comprehensive Education Data Model

One goal of this project was to determine the target audiences for a comprehensive education data model. However, before that could be determined, it was necessary to determine who is the target audience for the deliverables of *this* project (considered to be Phase I). The expert panel agreed that in addition to NCES, a target audience is the group of developers and data architects that will build out the model (in Phase II). The most difficult task in creating a comprehensive education data model is having a starting point that represents the best thinking to date. The categorization of the data sets and the comparison to the existing examples will give the architects and developers a powerful starting point to build out the section models that the states, LEAs and schools can utilize.

The target audiences for the comprehensive education data model include:

1. Vendors that help educational organizations build out their data systems. The modules and models that will be included in the comprehensive education data model will allow the vendors to ensure their products serve the core needs of their clients so that the



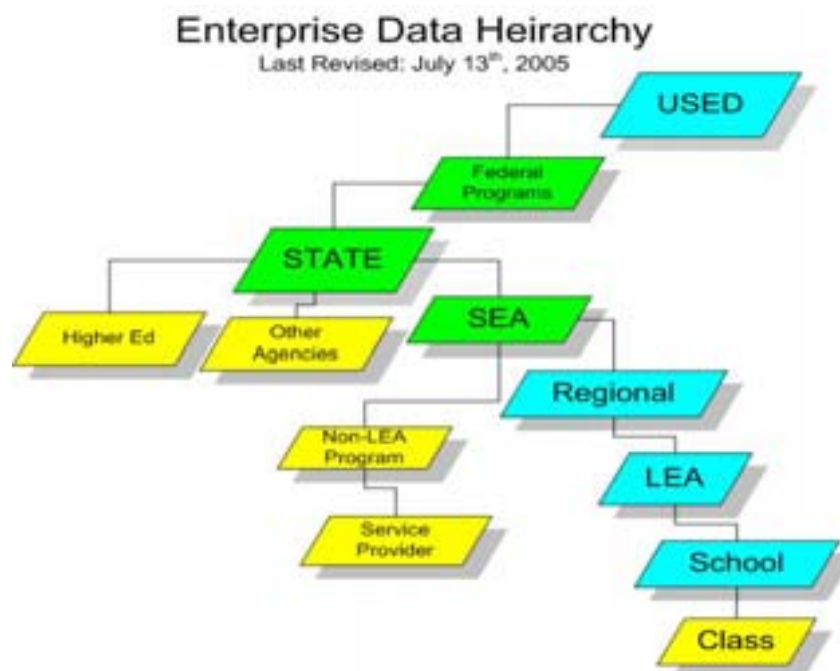
Comprehensive pK-12 Education Data Model

education organizations do not need to backtrack and do more work to cover basic requirements.

2. Technologists in the educational organizations that are designing and planning their data warehouses, data analysis, and decision support systems. This model will allow them to start from a baseline that covering 80% of the needed areas. These technologists can then focus their attention on customization and detail particular to their educational environment. This will save them time and money.
3. Business managers and administrators who are writing and evaluating Requests for Proposals (RFPs). Having a sample model laid out for them as an example to work from will help them to create more effective RFPs. It will also allow them to evaluate vendor responses more rigorously and actually get what they want. Although this is ranked as the third target audience, it may be the audience that has the biggest resource impact on education organizations in terms of saving time and money and moving states, LEAs, and schools forward quickly.

Identifying the Targeted Levels of the Model

Currently there are few education data models available below the state level - but the greatest need exists at lower levels. Since the data collected by state education agencies come from school districts (where much of the data begins) it is important to show the relationships among the levels. Because we know more about the top level, the project began at the top of the hierarchy and worked its way down. The goal is eventually to have the more detailed data at the school and district level. This drawing depicts the hierarchy of data flow as we have identified it for the purpose of moving data from the lowest level - the “class” - to the highest level - “federal”. Although there are relevant data collected by other entities, such as higher education and other state and local agencies, the focus of this model is pK-12.





Comprehensive pK-12 Education Data Model

Identifying Levels of Detail in the Model

Categorization of data can be at various levels of detail. The expert panel elected to begin looking at the NCES data hierarchy to determine how much to include in a high-level conceptualization of a pK-12 education data model. The NCES data hierarchy is as follows:

- **Domain** – The entity about which the data elements are collected. This is a way of grouping the listing of data elements into a subset for specific purposes. Student, staff, school, local education agency, intermediate education unit, and state education agency are the domains included.
- **Section** – A major high level grouping of data elements for a domain, such as personal information, educational experiences, assignment, etc. for staff.
- **Category** – A more refined grouping of data elements within a section.
- **Entity** – A person, place, event, object, or concept about which data can be collected.
- **Data Element** – A unit of data that can be defined and measured. Data elements may be open-ended or coded (having recommended responses or alternatives). If a code is warranted, a list of options is provided.

For the purpose of this project, work has concentrated on identifying the Domains, Sections and Categories that are represented in the data models that have been submitted for comparison. As these are moved into a true logical model, some levels may be collapsed or revised. There are numerous ways to represent the data, but the objectives for this project and the follow-on project is to pick the “best practice” levels that are useful.

Factoring the Domains, Categories and Sections

The identification of the most appropriate domains/categories/sections was a process of the following activities:

- Mapping to the categorization already done in the data model examples we have;
- Looking at the functional categorizations in the various Departments of Education and the EDEN domains; and
- Tying the categorizations to the NCES data structures.

It was logical to start with the NCES online handbook because it was the most complete framework starting point. However, that handbook does not identify program and class as domains, something the expert panel considered important.

These data categories, suggested by expert panelists, are very representative of the SEA viewpoint and represent data categorization by aggregate reporting and management needs as opposed to operational categories. These categories also matched most of the state data models collected for this project. This proposed model needs to reflect both kinds of factoring. The physical models reviewed from the states and districts represent this breakdown well, although sometimes in a slightly different approach.

DRAFT not for Dissemination

- | | |
|---|---|
| 1. Attendance | 11. Free and Reduced Price Lunch |
| 2. Annual Yearly Progress | 12. Graduates/Completers |
| 3. Courses | 13. NCLB (e.g. school choice, supplementary education services) |
| 4. Directory (schools and districts) | 14. Promotion/Retention |
| 5. Discipline | 15. Special Education |
| 6. Dropout | 16. Staff |
| 7. English Language Literacy | 17. Standards |
| 8. Enrollment | 18. Student Performance |
| 9. Federal Programs (e.g., after school, migrant) | 19. Vocational Technical Education |
| 10. Finance | |

Developing a Common Vocabulary

In order to ensure that all audiences can easily use the comprehensive education data model, it will be necessary to have a shared vocabulary and context for the terms and structures that are used. Existing lists of terms were identified and are included in Appendices C. This listing of terms and definitions will need to be revised in the follow-on project as the terms are used in the model.

Identifying Essential Data Model Components

The project expert panel identified the essential components to be included in the comprehensive education data model that should be developed in a proposed follow-on project. These include:

- Data Dictionary – All of the essential data elements should be identified and defined.
- Entity-relationship diagram (ERD) – A graphical representation of the essential data elements and entities and their relationships.
- Security/sensitivity documentation (FERPA, Policy, law, minimum n-value) – A discussion of the security and privacy requirements relating to the essential components of the data model.
- Time representation – A discussion of how granular data elements at one level relates to aggregate data elements at higher levels, as well as issues and/or changes needed at all levels.

IV. Project Findings

Summary of Data Model Comparisons

There are several driving conditions influencing the comparisons:

1. Most of the models were, or are close to being, physical data models. While this is appropriate for seeing what is actually being implemented at the state, district and local level, the particular database management systems and organization-specific workflow designs make it difficult to abstract the conceptual principals that gave rise to the model.



Comprehensive pk12 Education Data Model

2. It was discovered that many of the organizations never built a conceptual data model at all. They had their technologists build a logical data model with the physical model in mind.
3. Some of the “data models” are not relational models at all. They do not address the relationships between the elements; rather they are lists that include the narrative definition and the code values and possibly the physical field description of the elements. Essentially they are data dictionaries.
4. Due to the wide variety in the models and differences in scale, we compared the models at the level of section and category so as to have an “apples to apples” comparison.

The results of that comparison can be seen in two documents:

1. The two framework conceptual models, and
2. The comparison spreadsheet associated with this document (Appendix A)

Upon close comparison, the existing models varied greatly in design and function. While many were physical in nature, it was possible to draw from them the essence of data they were attempting to represent. While the NCES Handbooks represented the most descriptive data, it is not in the true form of a data model. Although it represents the most comprehensive elements for education currently available, the PBDMI / EDEN information is very comprehensive at the federal level but data is not found to be included in the SEA or LEA models. This suggests that the data elements at these levels would not be included if a comprehensive data model was attempted by an SEA or LEA. If they were included, this would streamline reporting and reduce the burden on schools and states.

While overlaps have been exposed, it is obvious the data represented by all the models varied significantly at all levels. The greatest amount of overlap was in the enrollment, program and assessment sections. One would think that the student and staff domains would be well represented in all of the models but, comparison of the overlaps revealed that this was not the case.

The largest gap that the expert panel encountered was the lack of details at the class level. They felt this gap has the biggest impact on education and learning. It exposed the data needs that are inherent if LEA and SEAs are to do precise and comprehensive longitudinal data studies. However, this level of detail would be a focus area that should be addressed in a follow-on activity. The panel recognized another gap as the need for relationships between the data and domains to be in a conceptual and logical manner. Without this detail in a comprehensive data model LEAs and SEAs may struggle to understand the data at its lowest level where it impacts students and staff.

Preliminary Data Model Sections

As a result of the review of the sample data models and discussion by the expert panel, preliminary data model sections were identified which were greatly influenced by the NCES Handbook organization. This was done to organize the information in a linear way and is considered “functional groups of information” (i.e. HR, finance, etc. information).



Comprehensive pk12 Education Data Model

Based on information in the Data Handbook (appendix B), there are four “organization” domains (SEA, IEU, LEA, School) and two “person” domains (Student, Staff). In this model, each of these domains would be core tables or groupings. There was some discussion about the use of Program and Class as “organization” domains. But the panel determined that that task would be for the follow-on project to decide. While these elements function much as a school functions in a data model, they are complex entities and may be best treated as functional domains.

Each of these NCES domains was further described by “categories/sections”, now called “Sections.”

Organization Domains

Inside the Organizational Domains there are nine sections identified as essential:

- Institutional Identification – general data about the name, address, affiliation, and functioning of an organizational entity
- Governance – data about the organization responsible for policy and operational decisions of an organizational entity and its rules and functioning.
- Programs – data about local, state or federal programs offered.
- Accountability – data about standards and performance.
- Instructional Programs – data about the instructional courses and programs.
- Activities – data about extra-curricular and co-curricular activities.
- Assessment – data about assessment given and results.
- Facilities – data about the physical structures in the education system.
- Safety and Discipline – data about crime, discipline, and safety issues.
- Finance – data about revenues, expenditures, assets, debts, etc.

Person Domains

The two person domains are quite different from each other.

Student

The Student domain has the following sections:

- Personal Information – data about a student’s personal, family, and demographic status.
- Enrollment (and exiting) – data about a student’s entrance, membership and exit from the education system.
- School Participation and Activities – data about a student’s classes, outcomes, and graduation.
- Non-School and Post-School Experiences – data about training received outside of school, work experience, honors received, and education or posts held after leaving high school.
- Assessment – data about assessment purpose, administration, response, and scoring.
- Transportation – data about moving student to and from school.
- Food Services – data about food materials ordering, management as well as student accounts management



Comprehensive pk12 Education Data Model

- Health Conditions – data about students’ physical condition, limitations, health history, and immunizations.
- Special Program Participation and Student Support – data about student participation in non-instructional programs.

Staff

The Staff domain has the following sections:

- Personal Information – data about a staff person’s personal, family, and demographic status.
- Educational Experiences – data about courses taken, degrees received, and other training.
- Qualification Information – data about credentials, prior related experience and other interests.
- Current Employment – data about conditions of employment, compensation, benefits, attendance, and grievances.
- Assignments – data about assignment, operational unit, schedule, workload, program area, and function.
- Evaluation and Career Development – data about quality of job performance and career development.
- Separation from Employment – data about conditions of termination.

The Developed Domain Framework

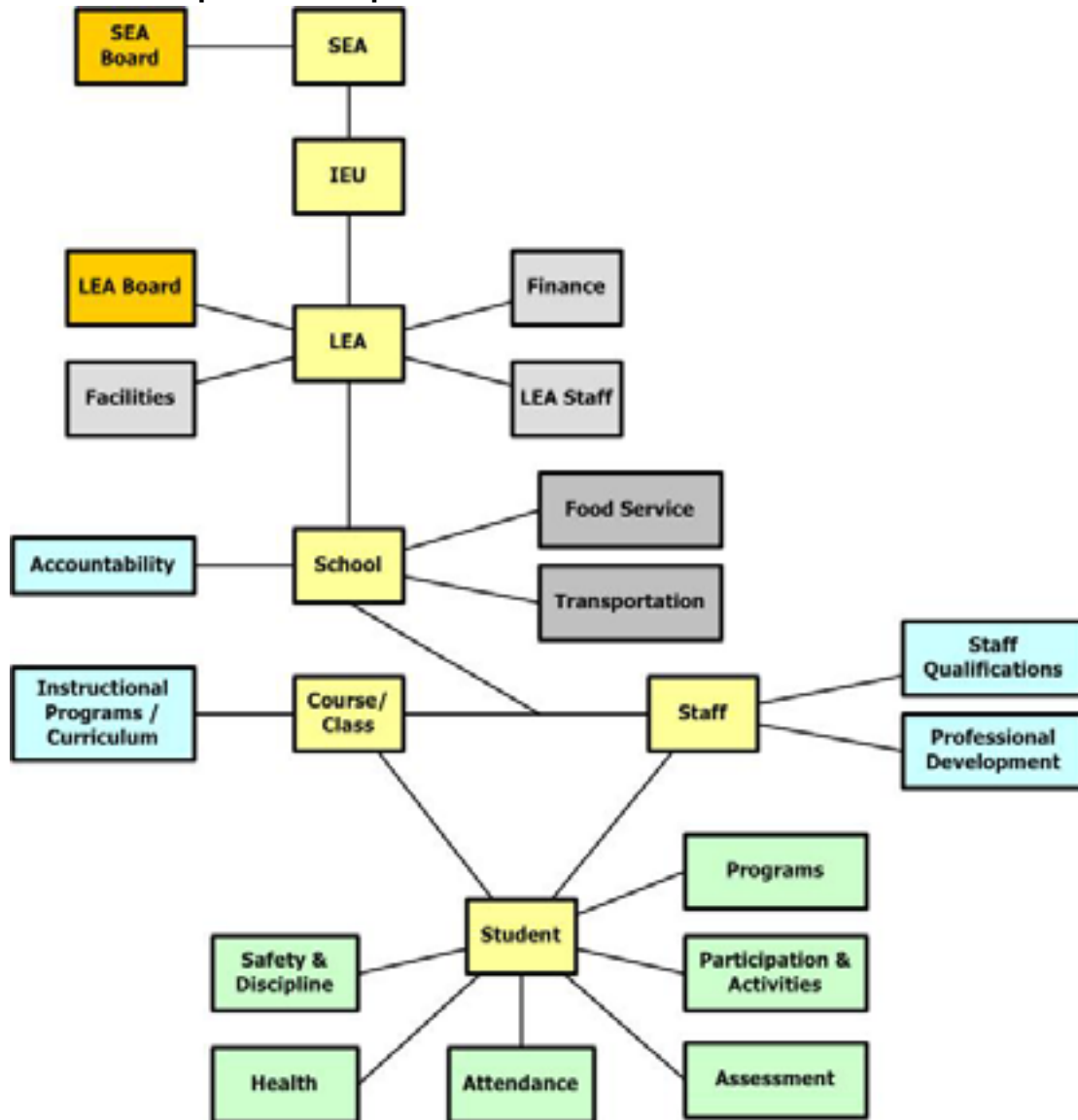
In response to the complexity of this previous model built off existing federal, SEA and LEA data model work, the expert panel generated a simplified diagram to represent a more accessible view of the data model categories and sections. In this diagram: yellow represents the major “ownership” institutions, gold is legislative or policy oversight entity, grey is various operational data areas, aqua is staff and instructional components and green is student specific data components.

This model does not represent the only “correct” way to structure a state, LEA or school data system. It is meant as a framework informed by existing models, in which the most common relationships and ties between data sets can be represented. As each organization adopted sections of the framework, it would need to make the definitive relationship decisions that would represent its unique data needs and utilization requirements.

This modularization and detailing for each of the sections will be the primary output of Phase II of this project. These features and will be the primary value-add to states, LEAs and schools struggling with the challenging prospect of creating a thorough and fully enabled data system. These sections (boxes) are represented below (i.e. finance box separated with the ten organization domains identified on page 19).



Comprehensive pk12 Education Data Model



V. Suggested Next Steps

With the lack of a comprehensive education data model available to states and schools today, the logical next step is for a large-scale stakeholder engagement strategy around the development of just such a model for federal, SEA and LEA utilization. This future deliverable would not only facilitate local decision-making but add to overall data quality by aligning data to be aggregated at the source. It would produce the “cleanest” data available. It is critical that this engagement would involve all relevant stakeholders in data utilization at the national, state, local and even the classroom level. This includes educators, policy makers, researchers and even consultants and vendors.

The next steps should include but are not limited to:



Comprehensive pk12 Education Data Model

A Comprehensive Education Data Model: There is a need for a logical education data model representing all of the relevant levels that exchange data. Among the important relationships that must be shown at each level are:

- One-to-one relationships.
- One-to-many relationships.
- Many-to-many relationships.

A typical ERD diagram can show these relationships and be considered a “best practice”.

Identified Data Flow: It is also important to show how data maintenance at the lowest level relates to how the data flows to higher levels in the model. Among the issues to be addressed are:

- Between which levels do the data flow?
- Who are the data stewards/owners?
- For whom/what purposes are the data?
- Who will be using the data and for what purpose?

This may be the most difficult part of the work, but it is the most important. Data that are maintained in individual student records are often aggregated as they flow up into higher levels. Ideally the data model would show how an individual characteristic in a person’s record would relate to a total of people with those characteristics.

Modular in design: Project participants recognize the need to split off sections of the data model to further explore relationships. For instance, the essential components of a model for a particular program may require greater detail than those in another program.

- The Comprehensive Education Data Model should provide resources that will allow people to quickly locate the data in a modular fashion.
- The model should allow a person to concentrate on a specific area without having to deal with the rest of the model.

Class domain: This area is the least well-described area in the models reviewed. This is probably due to the fact that instructional technology and other classroom-based activities were not typically a part of the overall education data system. This is changing, however. What happens in the classroom is under more scrutiny than ever; thus it deserves to be a critical component in the model.

- The classroom is the essence of teaching and learning and should be monitored to ensure accountability. This cannot be done without consideration in the data model.
- Impact on the staff and students is greatest at this level. The class serves as the nexus for students, staff, and instructional (and other types of) programs. More attention can only push the education community further in identifying ways to improve instruction.

Tightened Definitions: It will be critical to decide how to model certain things and how to give districts and states the best guidance:

- Should we describe logical entities from the viewpoint of the designer and policy maker or the implementer and end user? Examples of this include



Comprehensive pk12 Education Data Model

Curriculum, Discipline, and Transportation. Not only is the authority for these activities distributed differently district to district and state to state but the implementation of them also happens at different levels.

- Are staff decisions always made at the district level or at the school level? Different examples gave different answers. When we move to the next step we will need to both locate the answer in the model and provide documentation to guide districts and states in their design.

VI. Conclusion

It is clear from the work done by SIFA - with the help of the expert panel - that there is a real need for a comprehensive pK-12 education data model. It would also be critical that this model is kept up to date through well-defined periodic review and stakeholder input mechanisms. Such a model will help to ensure that data systems can meet the needs of educators, not the requirements of vendors. Educators are the essential reviewers and users of a data model. If the questions that need to be answered cannot be obtained by the relationships of the data, then the model or system is flawed. Vendors also benefit from the use of such a model because it will indicate important components for educators.

If a comprehensive data model is developed with the information gleaned from this report, another key component to its usage and implementation will be the documentation and dissemination of best practices usage and possible tools for model utilization. A true “best practice” data model should show how all essential questions can be answered through the use of data within the education system. This could be considered Phase Three in the successful development of this comprehensive pK-12 data model.

“In the end, the real line of accountability will be drawn when a district goes to use or implement the comprehensive data model” – a sentiment delivered by a member of the expert advisory panel.

DRAFT not for Dissemination

APPENDICES

Appendix A: Data Model Comparison Crosswalks

Summary of Data Model Comparisons																													
Table One																													
Organizational Domains		SIFA Data Model SEA IU LEA Schools		PDE/M / EDEN SEA IU LEA Schools		NCES Handbook Information SEA IU LEA Schools		NCES Common Core Data SEA IU LEA Schools		California Department of Education SEA IU LEA Schools		Virginia Department of Education SEA IU LEA Schools		Maine Department of Education SEA IU LEA Schools		Washington DC Public Schools SEA IU LEA Schools		Richmond Public Schools SEA IU LEA Schools											
• Institutional Identification		X		X	X	X	X	X	X	X		X		X	X														
• Governance							X	X	X	X				X			X												
• Programs				X	X	X	X	X	X	X		X		X	X	X		X											
• Accountability				X	X	X	X	X	X	X		X		X	X	X		X											
• Instructional Programs		X		X	X	X	X	X	X	X	X	X		X	X	X		X											
• Activities							X	X	X	X																			
• Assessment		X		X	X	X	X	X	X			X		X	X	X		X											
• Facilities		X					X	X				X		X	X		X												
• Safety and Discipline				X	X	X	X					X		X	X	X		X											
• Finance		X		X	X	X	X		X	X		X		X	X	X		X											
Table Two																													
Personal Domains		SIFA Data Model SEA IU LEA Schools		PDE/M / EDEN SEA IU LEA Schools		NCES Handbook Information SEA IU LEA Schools		NCES Common Core Data SEA IU LEA Schools		California Department of Education SEA IU LEA Schools		Virginia Department of Education SEA IU LEA Schools		Maine Department of Education SEA IU LEA Schools		Washington DC Public Schools SEA IU LEA Schools		Richmond Public Schools SEA IU LEA Schools											
Students		SIFA Data Model SEA IU LEA Schools		PDE/M / EDEN SEA IU LEA Schools		NCES Handbook Information SEA IU LEA Schools		NCES Common Core Data SEA IU LEA Schools		California Department of Education SEA IU LEA Schools		Virginia Department of Education SEA IU LEA Schools		Maine Department of Education SEA IU LEA Schools		Washington DC Public Schools SEA IU LEA Schools		Richmond Public Schools SEA IU LEA Schools											
• Personal Information		X		X	X	X	X	X	X	X		X		X		X		X											
• Enrollment		X		X	X	X	X	X	X	X		X		X		X		X											
• School Participation and Activities		X		X	X	X	X	X	X					X		X		X											
• Non-School and Post-School Experiences							X											X											
• Assessment		X		X	X	X	X					X		X		X		X											
• Transportation		X					X							X		X													
• Food Services		X		X	X	X	X		X							X													
• Health Conditions							X									X													
• Special Program Participation and Student Support				X	X	X	X		X	X		X		X		X													
Staff		SIFA Data Model SEA IU LEA Schools		PDE/M / EDEN SEA IU LEA Schools		NCES Handbook Information SEA IU LEA Schools		NCES Common Core Data SEA IU LEA Schools		California Department of Education SEA IU LEA Schools		Virginia Department of Education SEA IU LEA Schools		Maine Department of Education SEA IU LEA Schools		Washington DC Public Schools SEA IU LEA Schools		Richmond Public Schools SEA IU LEA Schools											
• Personal Information		X					X			X		X		X		X													
• Educational Experiences							X			X		X		X		X													
• Qualification Information		X		X	X	X	X			X		X		X		X													
• Current Employment		X					X			X		X				X													
• Assignments		X		X	X	X	X	X	X	X		X				X													
• Evaluation and Career Development		X					X							X		X													
• Separation from Employment							X									X													

DRAFT not for Dissemination

Appendix B: Data Handbook

NCES Handbooks Online - Table of Contents - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Mail Print Word PDF Help

Address <http://nces.ed.gov/programs/handbook/toc.asp> Go Links Norton AntiVirus

[site index](#) | [EQ.gov](#) | **NCES** National Center for Education Statistics | Search NCES

Handbooks Online

An online resource for standard education terms, definitions and classification codes.

[Home](#) | [About the Handbooks](#) | [How to Use](#) | [Search Options](#) | [Reports](#) | [FAQs](#) | [Appendices](#)

Table of Contents

Listed below are the top levels of the Handbooks Online. Click on a link below for further details on that part of Handbooks Online.

1. [Domain: School](#)
2. [Domain: Local Education Agency \(LEA\)](#)
3. [Domain: Intermediate Educational Unit \(IEU\)](#)
4. [Domain: State Education Agency \(SEA\)](#)
5. [Domain: Student](#)
6. [Domain: Staff](#)

Version 2.0

NCES Headlines

- [UPDATED DATA! Public School District Finance Peer Search](#)
- [NEW! Projections of Education Statistics to 2014](#)
- [NEW DATA! Compare Public Libraries](#)

[NCES Home](#) | [Publications](#) | [Surveys & Programs](#) | [Quick Tables & Figures](#) | [Data Tools](#)
[Search](#) | [Help](#) | [News Flash](#) | [NCES Staff](#) | [Contact NCES](#) | [Site Index](#)

National Center for Education Statistics
[Institute of Education Sciences, U.S. Dept. of Education](#)
(map) 1990 K Street, NW, Washington, DC 20006, USA, Phone: (202) 602-7300

Internet

DRAFT not for Dissemination

Appendix C: Data Handbook Entities

Accountability

A report that provides information about the performance of the education institution. Accountability tends to focus on AYP (Adequate Yearly Progress) calculations.

Activity

The co-curricular or extra-curricular activity or activities under the sponsorship and/or direction of the school (e.g., student organizations, intramural and interscholastic athletics, publications, band, orchestra, and service activities) in which the student participates.

Assessment

A tool, instrument, process, or exhibition composed of a systematic sampling of behavior for measuring a student's competence, knowledge, skills or behavior. An assessment can be used to measure differences in individuals or groups and changes in performance from one occasion to the next.

Assignment (Staff Assignment)

A specific group of activities for which a staff member has been given responsibility.

Attendance

The act and frequency in which an individual is present where the individual is assigned.

Class

A setting in which organized instruction of course content is provided to one or more students (including cross-age groupings) for a given period of time. (A course may be offered to more than one class.)

Instruction, provided by one or more teachers or other staff members, may be delivered in person or via a different medium. Classes that share space should be considered as separate classes if they function as separate units for more than 50 percent of the time.

Course

The organization of subject matter and related learning experiences provided for the instruction of students on a regular or systematic basis, usually for a predetermined period of time (e.g., a semester or two-week workshop) to an individual or group of students (e.g., a class).

Credentials

An active certificate, license, permit, or other documentation held by an individual that authorizes the holder to perform certain functions or to make certain claims about his or her competencies in employment or assignment. Credentials are issued by a state agency (or in some cases by other organizations), post-secondary institution, or professional association based on education and training completed, experience, assessment, background verification, and/or other requirements.

Curriculum

Instructional content, resources, and practices (e.g., courses) prepared for and offered to students.

Facility



Comprehensive pk12 Education Data Model

An installation in which preschool, elementary and/or secondary instruction is provided; includes all buildings, structures, and other stationary items that are located on a single site or on contiguous or adjacent sites and that are used for school activities.

A piece of land, a building site, a building, or part of a building owned by and/or used for activities of an organizational unit such as a school or system.

Finance

The organizational unit of the LEA which handles all financial disbursement.

Food Service

The organizational unit of the LEA that handles food. This includes data on recipes, ingredients, free and reduced lunch status of students, etc. There can be large amounts of data in this module including sophisticated budget and stock management processes.

Health

The organizational unit of the LEA and the School that handles health, the nurses and any medical programs in the district

Intermediate Educational Unit (IEU)

A regional, multi-services public agency authorized by State law to develop, manage, and provide services, programs, or other options support (e.g., construction, food services, technology services) to LEAs.

Instructional Programs /Instruction

Activities dealing directly with students and/or with improving the quality of student learning. Instruction may be provided for pupils in a school classroom, in another location such as a home or hospital, and other learning situations such as those involving co-curricular activities; it may also be provided through various media such as television, radio, telephone, and correspondence.

Local Education Agency (LEA)

An administrative unit at the local level which exists primarily to operate schools or to contract for educational services. These units may or may not be co-extensive with county, city, or town boundaries.

Non-School Experiences

Experiences the student has had outside the school environment considered note worthy and that the school system captures.

Post-School Experiences

Experiences that the student has had post-graduation that are collected from various organizations and post-secondary institution

Professional Development

A planned, structured process through which an individual improves his or her job-related knowledge, skills, or attitudes. Such a process is designed to enable a staff member to grow within a profession or organization or to attain an initial or additional credential.

Program



Comprehensive pk12 Education Data Model

A program is a set (plan) of activities and procedures designed to accomplish a predetermined objective or set of objectives. There are many types of programs. Some key ones include: ELL, the federal Title programs, and Vocational and Technical Education programs. Each will have its own sophisticated data model. A program is often a whole system in itself.

School Program

A program or service that supplements the regular school program or serves as early childhood education and care for children prior to reaching school-attendance age. These programs and services are often, but not solely, available through federal, state, or local agencies, non-profit organizations and/or other community-based organizations (or assistance provided by these entities). Services may be instructional or non-instructional in nature and may be provided within or outside of a school building.

Special Programs

Programs designed for either children with IEPs (Individual Education Plans) or 504 plans. This category also includes Student Support Services.

School

An institution that provides preschool, elementary and/or secondary instruction and may provide other education-related services to students; has one or more teachers; is located in one or more buildings; has an assigned administrator.

School Food

The entity in the school which handles the cafeteria and ensuring proper nutrition and sustenance is provided to the students.

Separation from Employment / Layoff

Separation of an employee from a position to which he or she was appointed as a result of abolition of a position, lack of work, or lack of funds; the employee may have recall rights for a certain period of time under certain conditions.

Staff Member

An individual who performs specified activities for any public or private education institution or agency that provides instructional and/or support services to students or staff at the early childhood level through high school completion. For example, this includes:

- 1) an "employee" who performs services under the direction of the employing institution or agency, is compensated for such services by the employer, and is eligible for employee benefits and wage or salary tax withholdings;
- 2) a "contractor" or "consultant" who performs services for an agreed upon fee, or an employee of a management service contracted to work on site;
- 3) a "volunteer" who performs services on a voluntary and uncompensated basis;
- 4) an in kind service provider; or
- 5) an independent contractor or businessperson working at a school site.

Core staff data will include their demographics and their educational experience, as well as their current employment and assignment data.

Staff Assignment

The data correlating a staff member to a LEA, School and possibly classes



Comprehensive pk12 Education Data Model

Staff Qualifications

The minimum educational, experience, and personal requirements which must be fulfilled by a person preliminary to appointment or promotion.

See CREDENTIALS.

State Education Agency (SEA)

The agency of the state charged with primary responsibility for coordinating and supervising public instruction, including the setting of standards for elementary and secondary instruction programs.

Student

An individual for whom instruction, services and/or care are provided in an early childhood, elementary or secondary educational program under the jurisdiction of a school, education agency, or other institution or program.

Student Data Collection

This is the group of reports that comprise the collection of information from the schools to the LEA and to the SEA regarding students. Includes Attendance, Enrollment, Programatic and Dropout data among others...

Student Demographics

Non-assessment or enrollment data about the student such as their free/reduced lunch status, special education status, local ID, middle name, date of birth etc.

State Education Agency

An administrative agency (e.g. state or local education agency) responsible for providing or administering early childhood, elementary- and/or secondary-level instruction or educational support services.